

Incorporating Measures of Student Growth in Educator Evaluation



SCOTT MARION
CENTER FOR ASSESSMENT

MARCH 8, 2012

Overview of Presentation

2

- Review of Student Growth Percentiles (SGP)
- Use of SGP in school and educator accountability
- Discussion of Shared Attribution

Describing Student Longitudinal Growth

3

- Measuring student growth in academic achievement across or even within years is gaining popularity in education for many good reasons:
 - Students all start at different places and measuring change from where they start seems more fair to students and adults
 - “Status” (i.e., single point in time) results are very strongly related to non-school factors such as income and class
- Even though many people want to measure growth, doing so well is not as easy as it might seem.

Describing Student Longitudinal Growth

4

- There are many approaches for describing student longitudinal growth, ranging from quite simple to incredibly complex
- Simple sounds good, but is rarely valid
- Complex sounds worrisome, but is often necessary
- Goal is to find a model that is only as complex as necessary
- The Grading Schools workgroup evaluated several types of growth models and settled on Student Growth Percentiles

A Brief Review

5

- The advisory groups considered two main classes of models:
 - Value Added Models (VAM)
 - Student Growth Percentiles (SGP, also known as Colorado Growth Model)
- Importantly, neither of these are true “growth” models. They both describe change in scores relative to some expectation (keep this in mind)
- I know some of the following is redundant, but I also know it takes hearing/discussing these concepts several times to feel comfortable explaining it to others

Value-Added

6

- There is not universal agreement on the term.
- It is frequently used to describe the practice of including certain variables in a regression model in an attempt to isolate the teacher or school's contribution to achievement.
- These are typically variables such as demographics (e.g. poverty status) or prior achievement (e.g. a previous test score).
- Generally, the aim is to produce a growth expectation – and the difference between the actual growth and this expectation is considered the value-added component.

Student Growth Percentiles

7

- What does it mean to say that a girl grew 2.5 inches from ages 5 to 6? Is that typical, a lot or not enough?
- What does it mean to say that a student's score increased by 10 points in math from 4th grade to 5th grade?
- Would it help us to know that on average, girls grew 3 inches from 5 to 6 years of age and that 85% or so grew between 2 and 4 inches? What would we say about growth of 2.5 inches?
- Similarly, would it help us to know that on average student scores increased 7 points from 4th to 5th grade?
- We would have even more information if we knew that students scoring 155 in 4th grade had an average score of 162 in 5th grade?

Student Growth Percentiles

8

- Student Growth Percentiles (SGP) is a regression based measure of growth that works by evaluating current achievement based on prior achievement and describing performance relative to other students with the “same” prior achievement histories.
- This provides a familiar basis to interpret performance – the percentile, which indicates the probability of that outcome given the student’s starting point.
- This can be used to gauge whether or not the student’s growth was atypically high or low

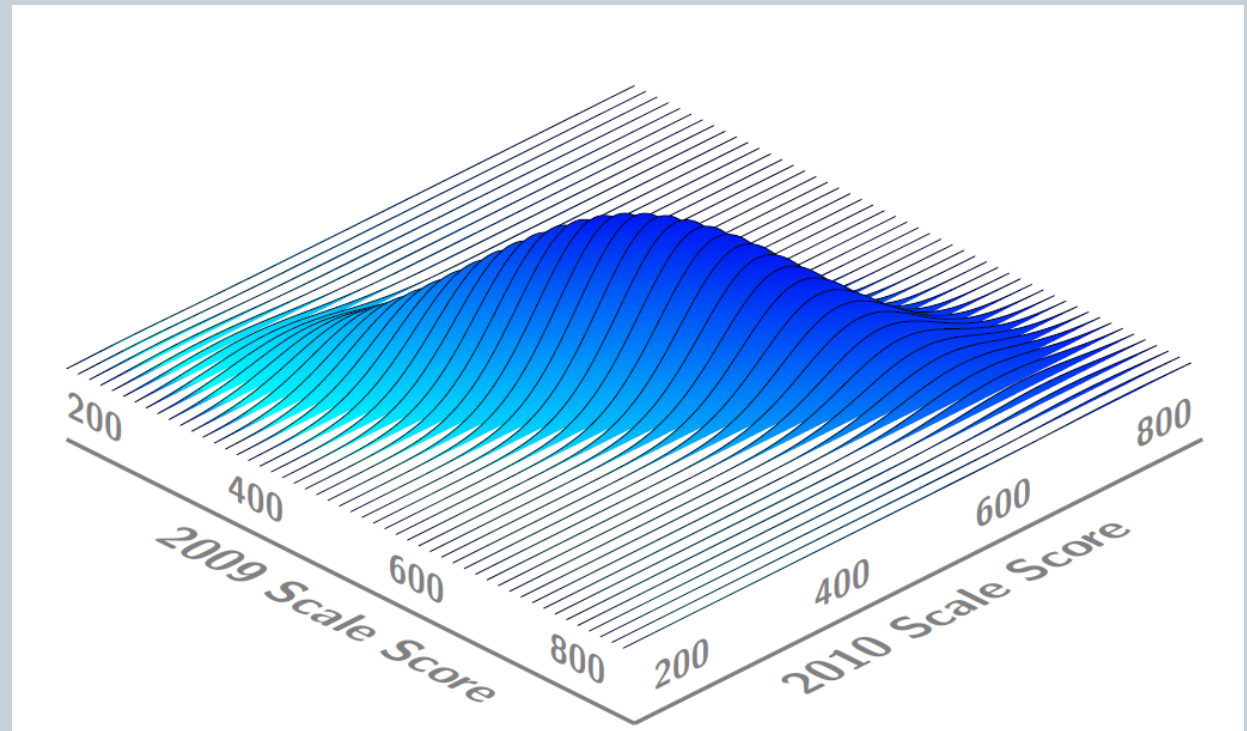
The Binominal Distribution

9

How does it work?

Think of a group of students, where each student has two test scores – one for 2009 and one for 2010.

We could show the distribution of these scores at the same time as pictured.

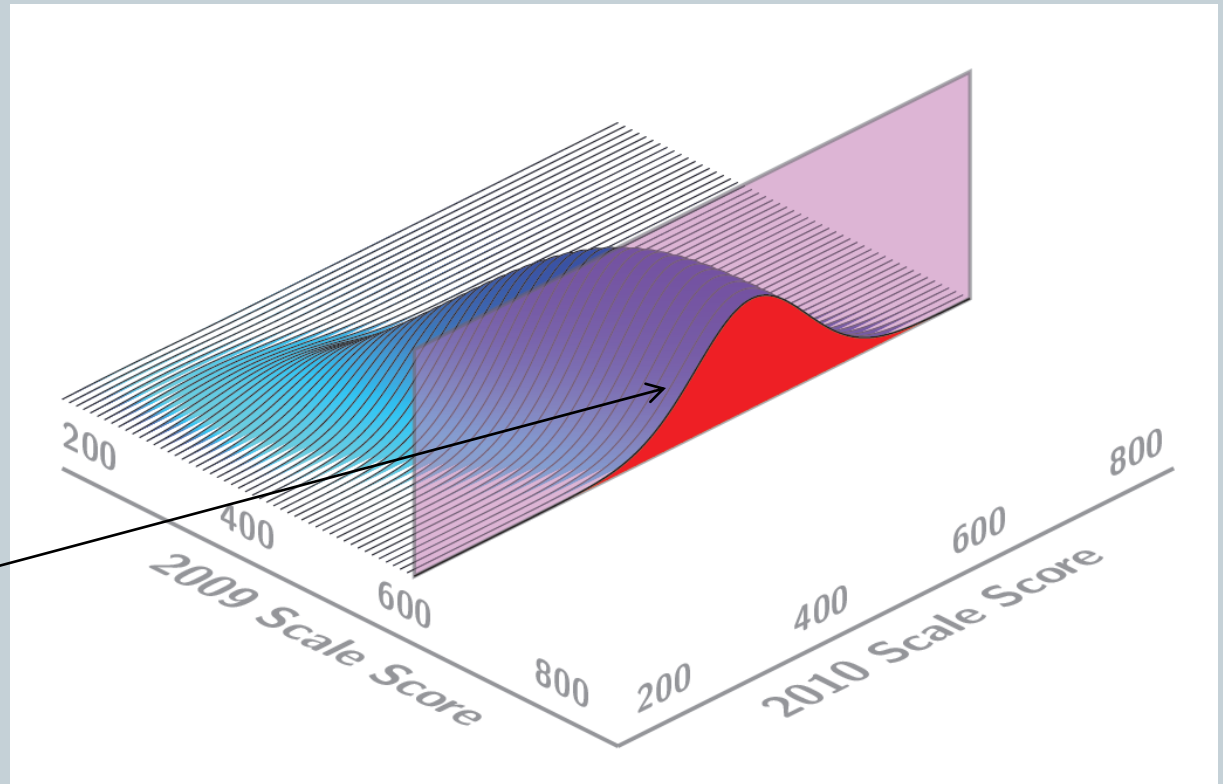


Slicing the distribution at the Year 1 score

10

We could 'slice' through the picture to show the 2010 distribution for just one 2009 score. This is called a conditional distribution.

The red shaded curve shows the conditional distribution in 2010 for all students who scored 600 in 2009.



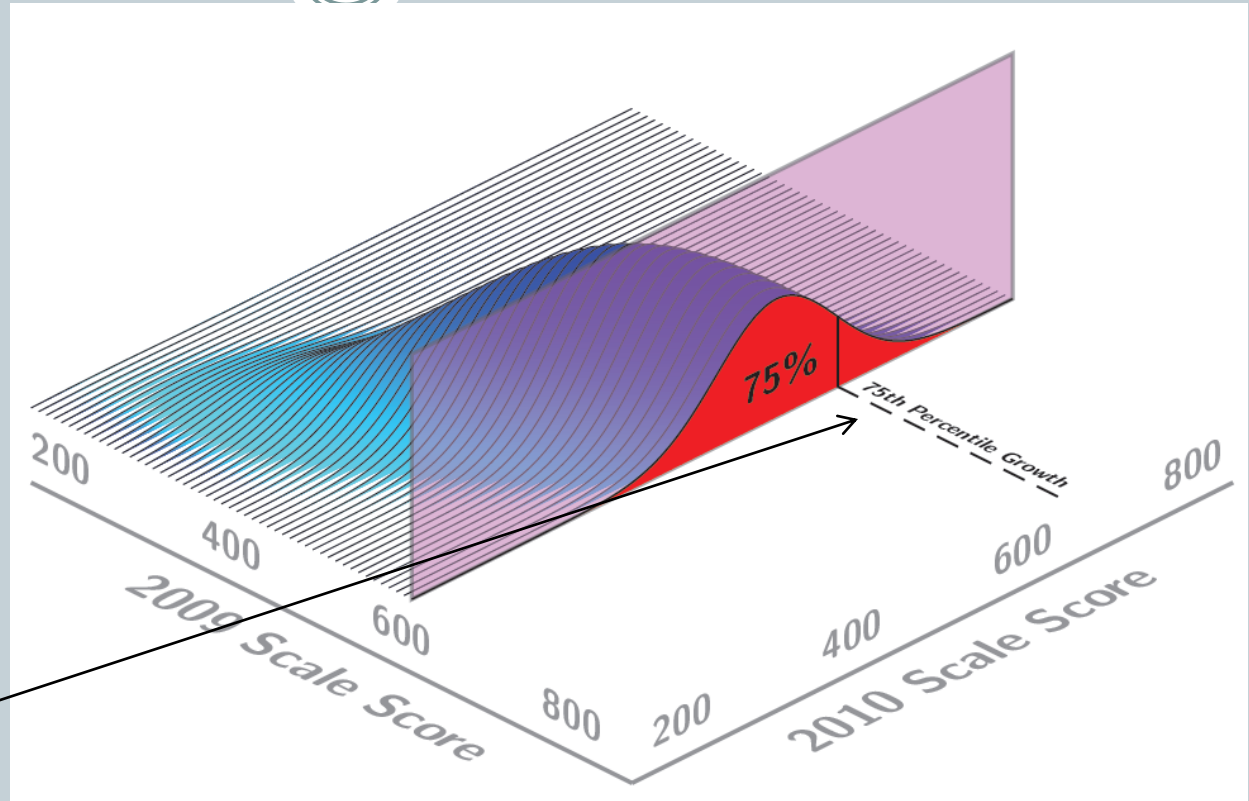
Comparing Year 2 scores for all who scored 600 in Year 1

11

Assume we are interested in just one score, 650, in 2010.

We could ask, what percentage of students who scored 600 in 2009 scored at or below a 650 in 2010?

In this case, that turns out to be 75%. In other words, a score of 650 is at the 75th percentile.

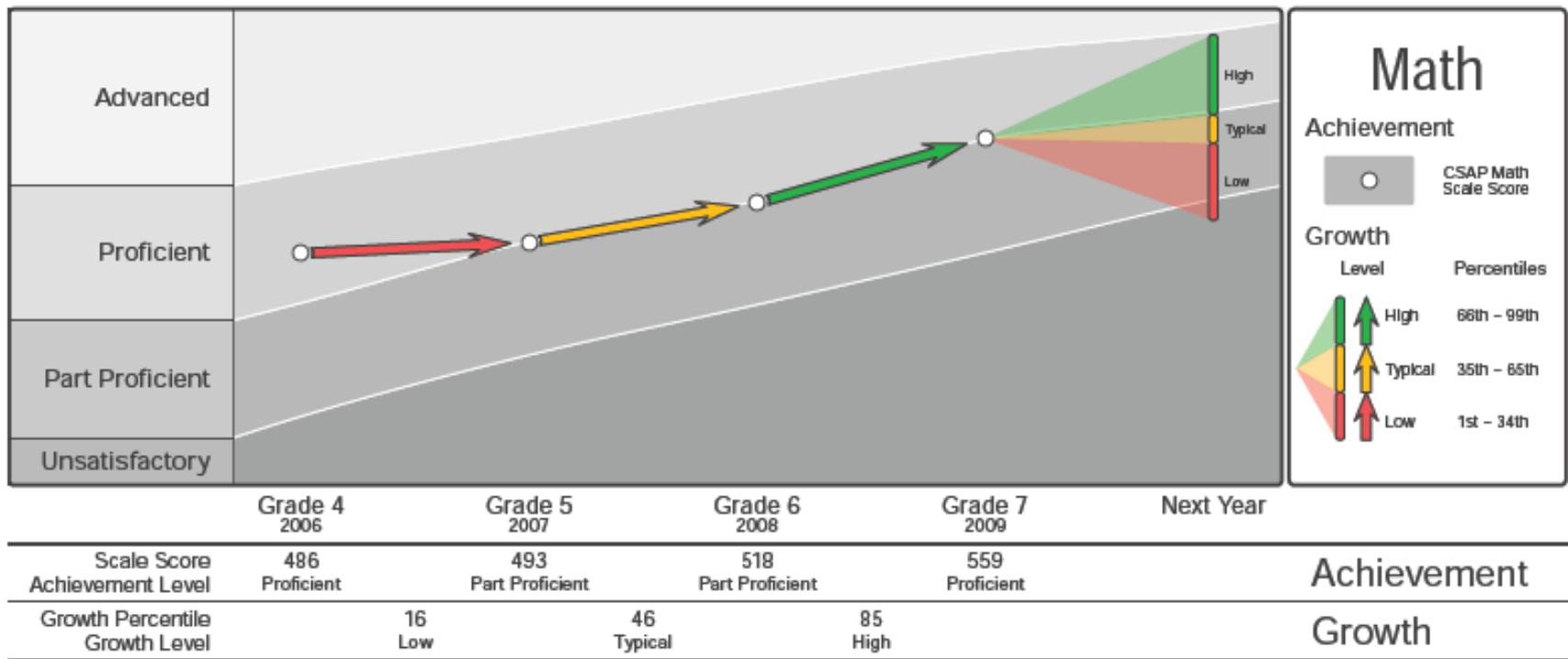


High level comparison of VAM & SGP

	VAM	SGP
Key Inference	Did students associated with teacher/school perform better/worse than prediction?	How does performance compare to students with same academic history?
Model “intent”	Explain variability, “statistical control,” “causation”	Describe likelihood of scores given score history
Unit Inference	Usually class or school, can extend to student	Student, but aggregate to any level (class, group, school, etc)
Conditioning	Prior scores, often demographics	Prior scores only
Metric	Normative: Residualized gain score	Normative: Growth Percentile
Types	Many variations including EVASS, VARC, etc.	Colorado Growth Model, but others (e.g., NWEA) use analogous approach
Proprietary	Generally yes	CGM=no; NWEA=yes
Complexity	Quite complex to extremely complex	Quite complex, but open source
Technical focus	Aims to minimize “bias” and increase precision	Hard to quantify “bias”, focus is on increasing utility

SGP: It all starts from the individual student

13



Aggregating student growth percentiles

14

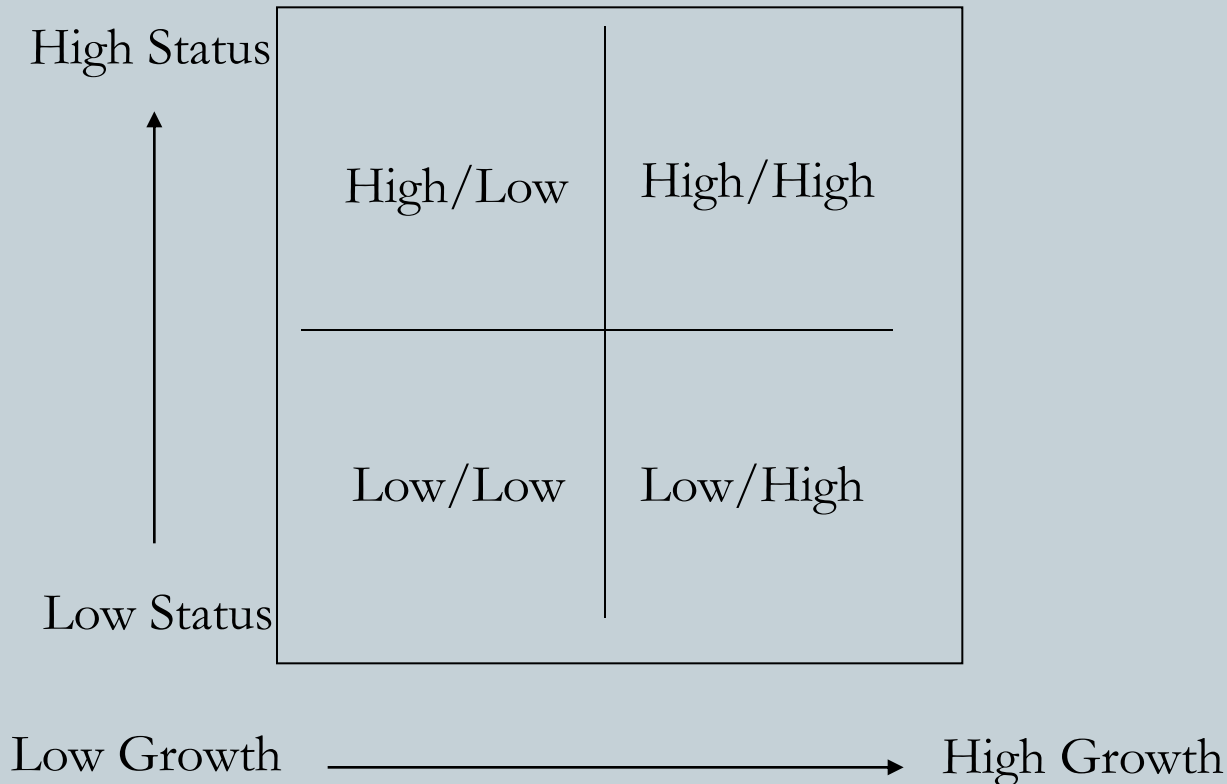
- Remember, we are developing educator and school accountability systems
- We need to be able to aggregate the individual student growth results to a class or school level
- We have found it most promising to use medians as the “average” school growth percentile
 - The median is the middle score in a distribution (the 50th percentile)
- For school accountability, we have also found it useful to consider the school’s achievement (status) in addition to growth

Growth and Achievement

15

A thoughtful approach to school accountability should consider how growth interacts with achievement (status).

Status/Growth Combinations

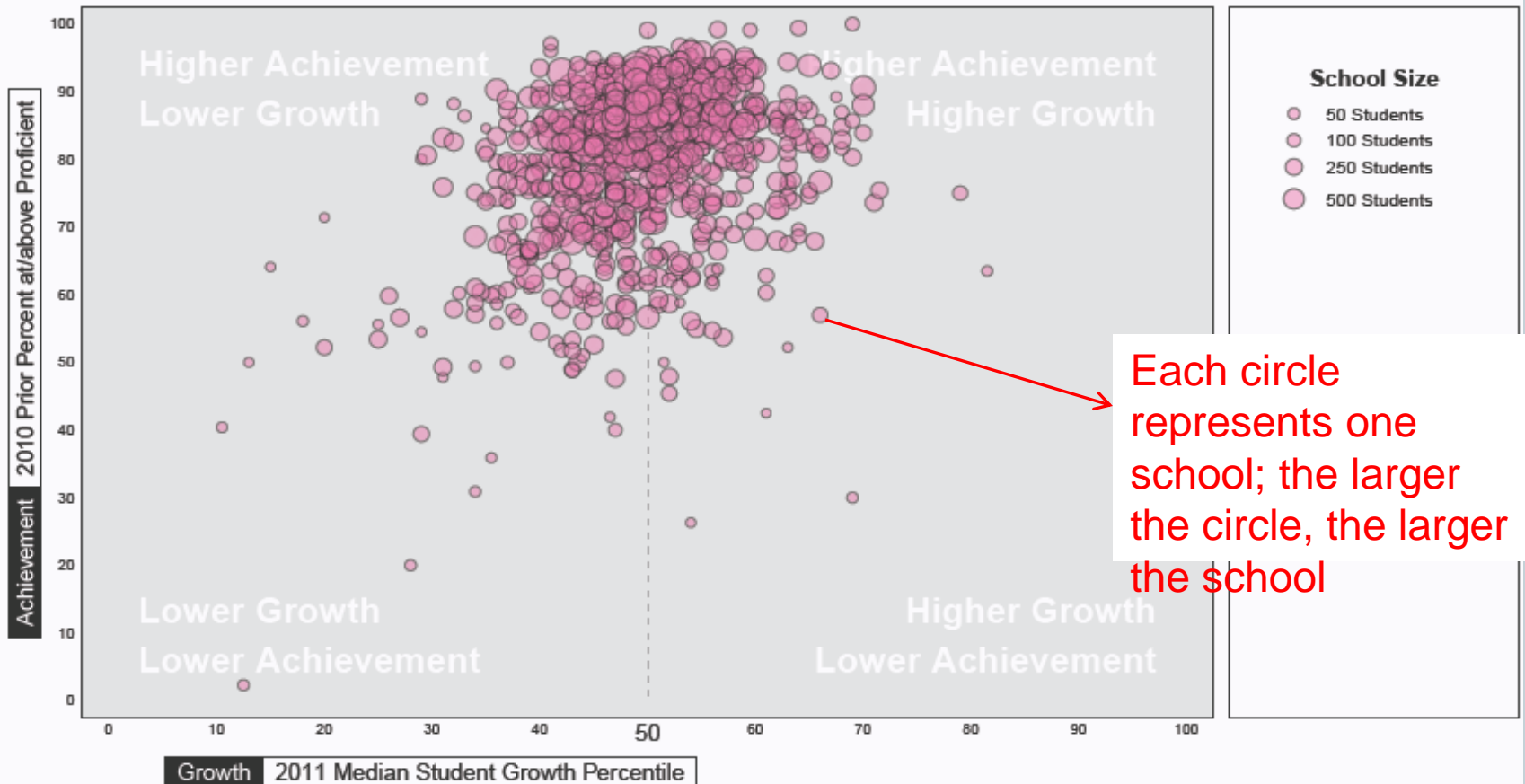


Relationship between Prior Achievement and Growth – ELA (All Schools)

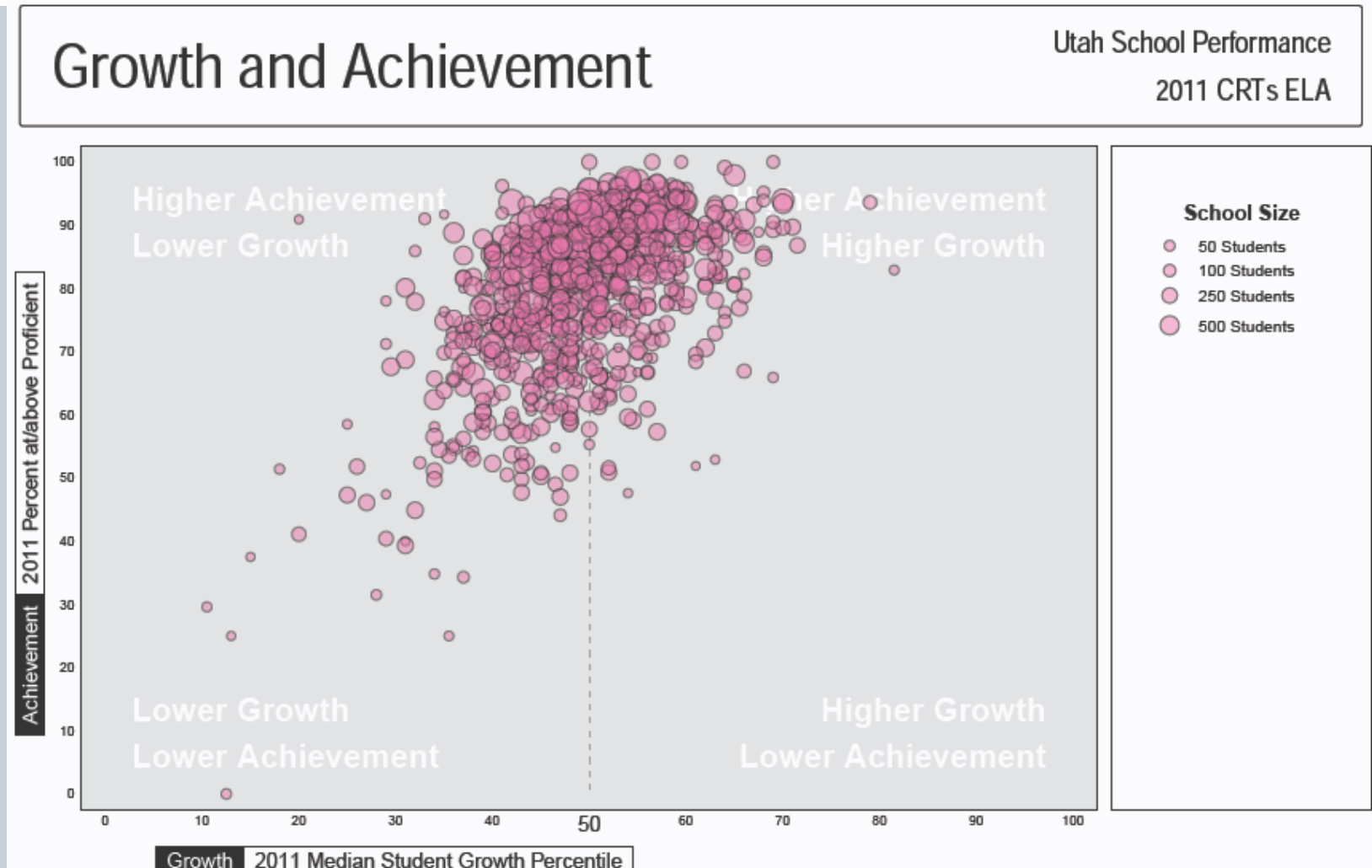
16

Growth and Prior Achievement

Utah School Performance
2011 CRTs ELA



Relationship between Current Achievement and Growth – ELA (All Schools)



Relationship between Prior Achievement and Growth – Math (All Schools)

(18)

Growth and Prior Achievement

Utah School Performance
2011 CRTs Mathematics

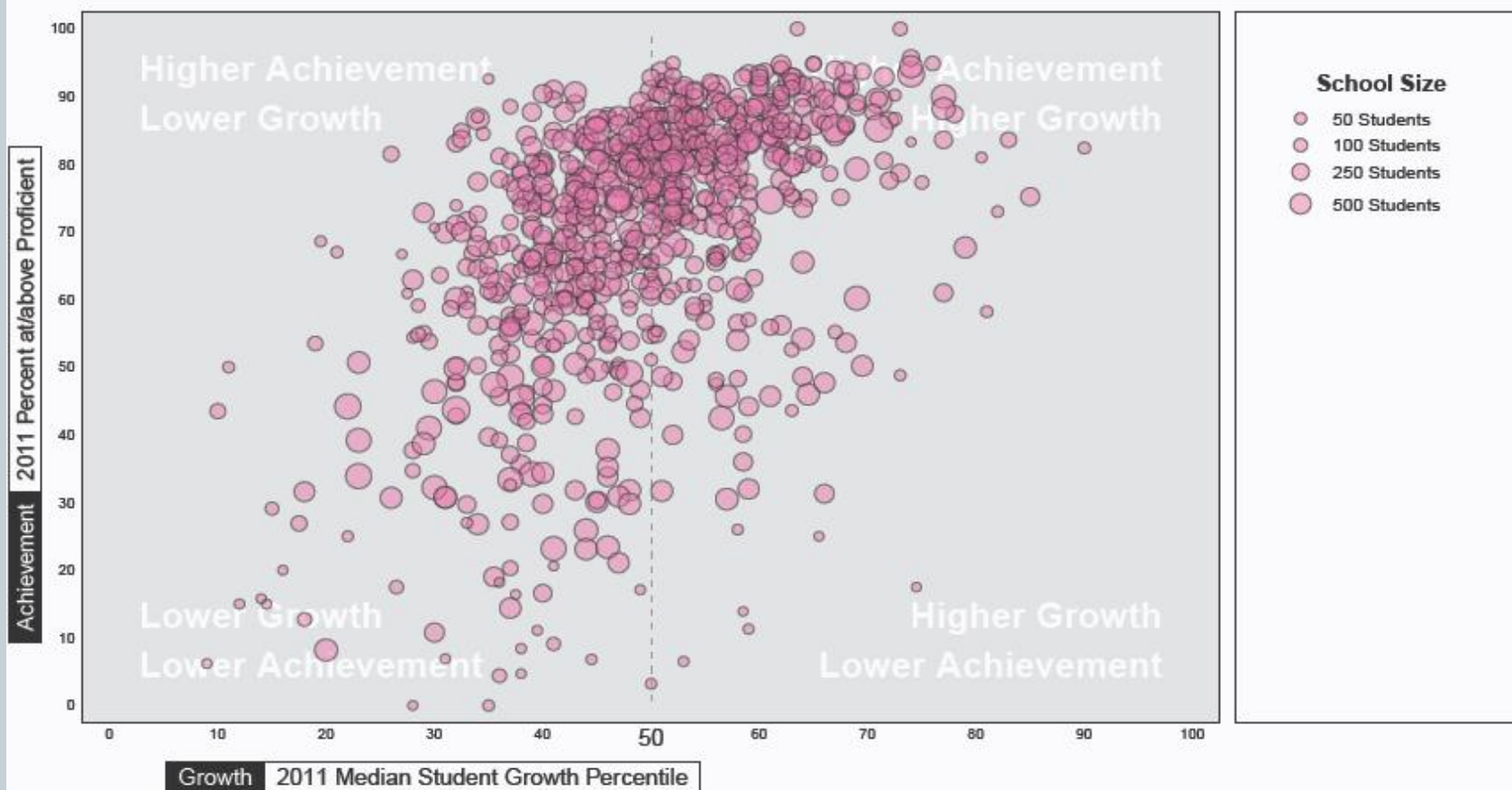


Relationship between Current Achievement and Growth – Math (All Schools)

(19)

Growth and Achievement

Utah School Performance
2011 CRTs Mathematics

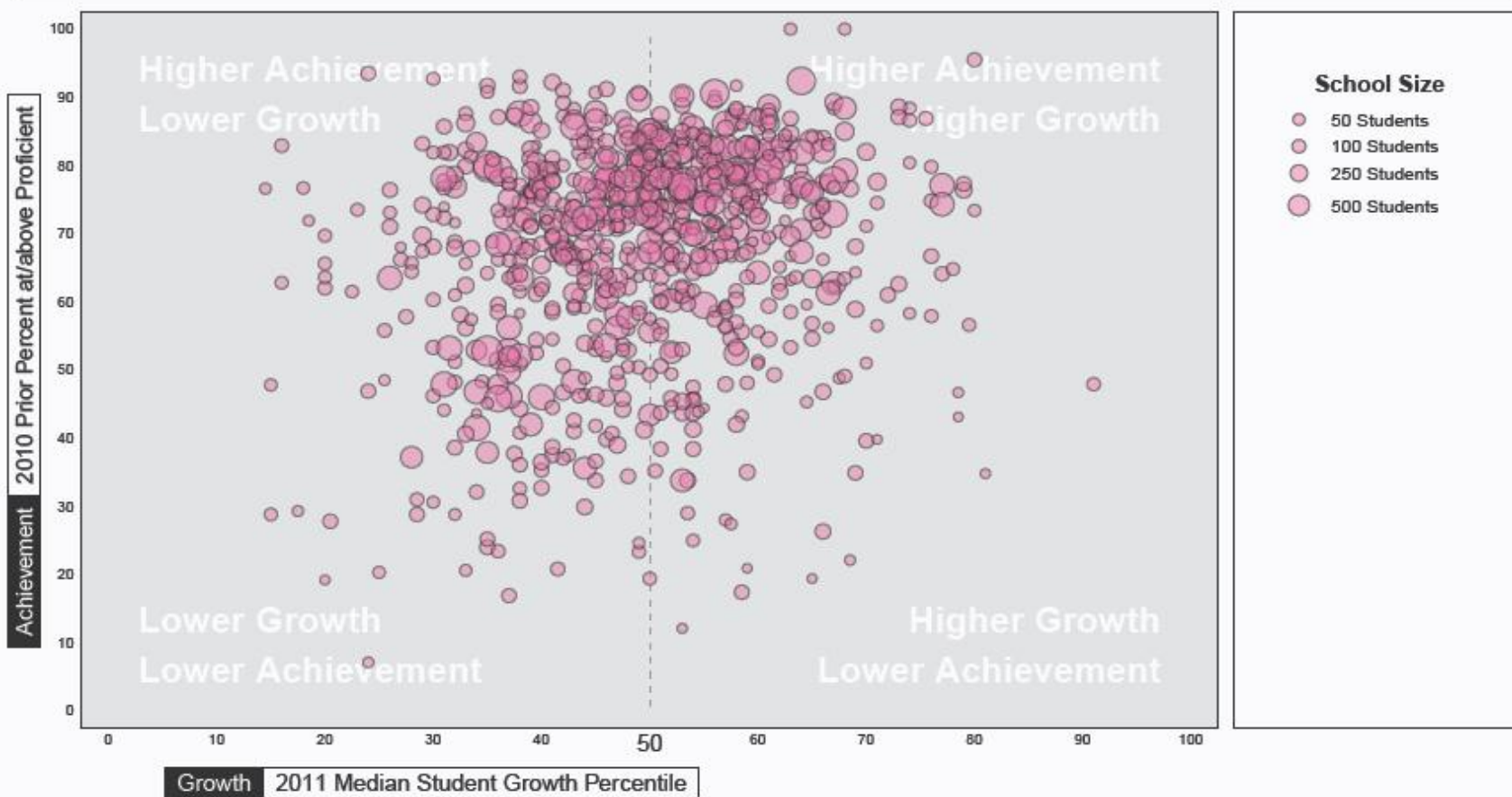


Relationship between Prior Achievement and Growth – Science (All Schools)

(20)

Growth and Prior Achievement

Utah School Performance
2011 CRTs Science



Relationship between Current Achievement and Growth – Science (All Schools)

Growth and Achievement

Utah School Performance
2011 CRTs Science



Growth for Educator Accountability

22

- While criterion-based growth can be very important for school accountability—although it is not part of Utah’s Comprehensive Accountability System—we are very concerned that it is not fair to base educator accountability on criterion-based growth
 - Highly correlated with socioeconomic status
- Therefore, we recommend using normative information for educator evaluations because it is more fair to all educators than a criterion-based approach

How many categories

23

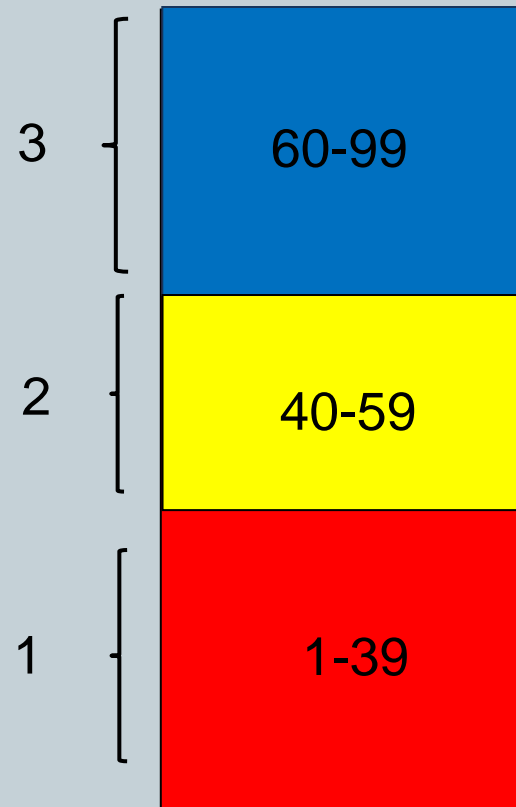
- Most states using SGPs (or VAM) for educator evaluations are categorizing growth into three categories:
 - High
 - Typical/Average
 - Low
- Why not more?
- Given the number of students included in SGP calculations for each teacher, it is doubtful that we can reliably distinguish among more than these categories

An Example of Potential Categories of Median Growth Percentiles (MGP)

24

- The specific median SGP cuts will have to depend on empirical analyses, but several states are using:
 - $MGP < 40$ = Low
 - $40 < MGP < 60$ = Typical
 - $MGP > 60$ = High

- Potential MGP rubric for educator evaluation



Shared Attribution

25

Shared Attribution

26

- Is the approach where median SGP or other (e.g., SLO) results are “shared” among more than the educator most closely associated with the SGP results
- Can be shared among all educators:
 - In the school
 - At a grade level
 - In a content area grouping (e.g., math department)
 - Other?

Theory of Action/Improvement

27

- Shared attribution should be based on more than just reliability concerns, but should be tied to your theory of improvement
 - For example, if the focus of improvement activities is the grade level team, that suggests attribution should be shared among educators at that grade
- What is your school's locus of improvement actions?
- Which subjects are shared and with whom? Does the team share both math and ELA results or just one subject?

Small group exercise

28

- Try articulate what types of shared attribution might make sense in your school. Consider the theory of action/improvement operating in your setting.
- What do you consider the advantages and disadvantages of shared attribution? Please try to be as specific as possible.

Tradeoffs of Shared Attribution

29

Advantages

- Larger sample sizes can lead to more reliable inferences
- Promotes collaboration among colleagues
- Avoids “isolating” or creating a hyper-focus on reading and math teachers

Disadvantages

- Educators are held accountable for results for which they may have little to no control
- Masks true variability in educator quality

Student Growth/Performance

30

- Now that we've talked about SLOs, SGP, and shared attribution...
- We need to create an initial plan for how we will incorporate measures of student performance in educator evaluations

Discussion

31

- What approaches (SLO, SGP, shared attribution) for documenting student performance (growth) will be included in the model system?
- Will the same approach apply to all educators?
 - If so, how shall we wrestle with the different sources of data?
 - If not, how will we ensure fairness among educators in the same school/district?
- We need to figure out how to put these various sources of evidence together in educator evaluations and consider the weight student performance will receive in the overall educator evaluation.